REMARKS/ARGUMENTS

The remarks are in response to the Final Office Action dated August 9, 2005. Claims 1, 2, 4-10 and 12-23 are pending in the present application.

Claim Rejections under 35 U.S.C. §112

The Examiner has rejected each of the pending claims under 35 U.S.C. §112, 2nd paragraph, for failing to particularly point out and distinctly claim the subject matter of the invention. The Examiner states, "[r]egarding independent claims 1, 9 and 17, [t]here is no criteria for retrieving a document to be indexed and temporarily storing in a storage device and it is not clear how information is determined to be relevant of not." Applicant respectfully submits that the claims are to be interpreted in light of the Specification. The Specification indicates that the present invention is directed to an information retrieval system that gathers documents from document repositories coupled to a network (Spec., page 8, lines 11 et seq.) and stores those documents temporarily in a storage device. Current search engines perform these same things. There is no criteria for which documents are retrieved because all documents that can be potentially indexed are retrieved and stored temporarily. According to the present invention, after a document to be indexed is retrieved, the extractor 209 analyzes the contents of the document using well known techniques in the area of data mining and sets threshold values to determine whether the document contains relevant information (Specification, page 9 line 19 to page 14, line 15). Accordingly, Applicant respectfully submits that independent claims 1, 9 and 17 satisfy the requirements for definiteness under 35 U.S.C. §112, 2nd paragraph.

Claim Rejections under 35 U.S.C. §103

The Examiner rejected claims 1, 2, 4-10 and 12-23 under 35 U.S.C. §103(a) as being unpatentable over Meyerzon et al. (U.S. Patent No. 6,631,369) in view of Nelson et al. (U.S.

Patent No. 6,243,713) and further in view of Matusbayashi et al. (U.S. Patent No. 6,473,754).

In rejecting the independent claims, the Examiner stated:

> Regarding claims 1 and 9, Meyerzon discloses a method for retrieving information using a search engine comprising the steps of:
>
> retrieving a document to be indexed and temporarily storing the document in a storage device (see col. 4, lines 43-54, Meyerzon);
>
> determining whether relevant information is contained in the document (summary, Meyerzon);
>
> generating a document extract corresponding to the document (see col. 4, lines 53-67); and
>
> storing the plurality of tokens in a search index, wherein the search engine accesses the search index to retrieve information in one or more document extracts satisfying a search query (see col. 7, lines 44-65 and col. 8, lines 1-10, Meyerzon. The data type of information corresponding to the "token").
>
> Meyerzon, however, does not explicitly disclose extracting a portion of the document that characterizes the document's subject content to form the document extract and decomposing the document extract. Nelson, on the other hand, discloses the retrieval system for retrieval of multimedia information including the extracting a portion of the document and decomposing the document into a plurality of tokens (see abstract of Nelson; col. 5, line 52-col. 6, line 65, col. 7, lines 46-67 and col. 9, lines 60-65). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify Meyerzon to include the claimed feature as taught by Nelson . . . .
>
> Meyerson and Nelson combination does not disclose "replacing the document in the storage device with the document extract." Matusbayashi discloses a method and system for extracting characteristic string and searching for relevant document including replacing the document in the storage devie with the text extract (see summary and col. 24, lines 19-25, Matusbayashi). . . .
>
> Regarding claim 17, Meyerzon discloses a system for retrieving information, wherein the system includes a search engine comprising:
>
> - means for retrieving a document from a documentary repository (see col. 4, lines 43-54 and element 200, Fig. 2 and corresponding text, Meyerzon);
>
> - an information extractor coupled to the means for retrieving, wherein the information extractor determining whether relevant information is contained in the document (summary, Meyerson), generates a document extract corresponding to the document (see col. 4, lines 53-67, Meyerzon). Each document is retrieved from the web site process and the data are extracted from each of these retrieved documents. Therefore, there must be an extractor for the extracting process;
>
> - a storage device (100, Fig. 2 and corresponding text, Meyerzon) coupled to the information extractor for storing the document extract;
>
> - a search engine indexer (300, Fig. 2) coupled to the storage device; and
>
> - a search index (400, Fig. 2) coupled to the search engine indexer for storing the plurality of tokens, wherein the search engine accesses the search index to retrieve information in one or more document extracts satisfying a search

query (see col. 7, lines 44-65 and col. 8, lines 1-10; Fig. 2 and corresponding text, Meyerzon).

Meyerzon, however, does not explicitly disclose the steps of extracting a portion of the document that characterizes the document's subject content to form the document extract and decomposing the document extract into a plurality of tokens. Nelson, on the other hand, discloses the retrieval system for retrieval of multimedia information including the decomposing the document into a plurality of tokens (see abstract of Nelson; col. 5, line 52-col. 6, line 65, col. 7, lines 46-67 and col. 9, lines 60-65). It would have been obvious to one of ordinary skill in the art at the time of the invention to modify Meyerzon to include the claimed feature as taught by Nelson . . . .

Meyerson and Nelson combination does not disclose "replacing the document in the storage device with the document extract." Matusbayashi discloses a method and system for extracting characteristic string and searching for relevant document including replacing the document in the storage devie with the text extract (see summary and col. 24, lines 19-25, Matusbayashi). . . .

Applicant respectfully disagrees.

The present invention, as recited in claim 1 provides:

> 1. A method for retrieving information using a search engine comprising the steps of:
> (a)   retrieving a document to be indexed and temporarily storing the document in a storage device;
> (b)   determining whether relevant information is contained in the document;
> (c)   if the document contains relevant information, generating a document extract corresponding to the document by extracting a portion of the document that characterizes the document's subject content to form the document extract;
> (d)   replacing the document in the storage device with the document extract;
> (e)   decomposing the document extract into a plurality of tokens; and
> (f)   storing the plurality of tokens in a search index, wherein the search engine accesses the search index to retrieve information in one or more document extracts satisfying a search query.

Independent claims 9 and 17 are computer readable medium and system claims, respectively, having scopes similar to that of claim 1.

<u>Independent claims 1, 9 and 17 are Allowable.</u>

Applicant respectfully submits that none of the cited references, alone or in combination, teach or suggest the cooperation of elements recited in claims 1, 9 and 17. In particular, none of the references teaches or suggests determining whether relevant information is contained in the document and generating a document extract if the document contains relevant information, extracting a portion of the document that characterizes the document's subject content to form the document extract, and replacing the document in the storage device with the document extract, as recited in claims 1, 9 and 17. In the presnet invention, a document extract is generated if the document contains relevant information. Moreover, a search index is based on the document extracts which characterize the subject content of the corresponding documents. Accordingly, the search index is based on the *semantic value* of the documents, as opposed to just the words or components of the document.

In contrast, Meyerzon is directed to minimizing the number of requests a web crawler makes to a document server to obtain the "increment" of the document set relative to the set of documents received during the previous crawl. Nelson is directed to indexing compound documents in a unified common index. In Nelson, a compound document, i.e., a document containing multimedia components, is broken up into its constituent components (e.g., text, audio, images) and one or more tokens is created for each component. The components and their tokens are then stored in the unified common index (col. 2, lines 19-27).

Matsubayashi is directed to extracting features in contents of a seed document without using a word dictionary and a system using the extracted features to search for other documents related to the seed document. In Matsubayashi, single character type strings of single character types are extracted and an occurrence frequency is calculated. This is repeated for the other

documents in the repository. Documents related to the seed document are found by comparing the occurrence frequencies associated with the single character type strings. (Summary).

While Meyerzon teaches "extracting the data from each of these retrieved documents and storing the data in an index" (column 4, lines 55-59), and Nelson teaches decomposing the compound document into its constituent multimedia components, indexing the components, and storing the indexed data in an index (column 5, lines 52-67), neither reference focuses on building an index based on *relevant* documents and on the documents' *subject content*. In particular, neither Meyerzon nor Nelson, singularly or in combination, teach or suggest "determining whether relevant information is contained in the document," and "if the document contains relevant information, generating a document extract corresponding to the document by extracting *a portion of the document that characterizes the document's subject content* to form the document extract" corresponding to the document. Moreover, the combination of Meyeraon, Nelson and Matusbayashi do not teach or suggest "replacing the document in the storage device with the document extract, " as recited in claims 1, 9 and 17.

In the Final Office Action, the Examiner states that the "summary" of Meyerzon teaches determining whether relevant information is contained in the document. Applicant respectfully submits, however, that Meyerzon teaches no such thing in the summary. Rather, Meyerzon merely discusses how the web crawler can avoid checking the time stamp for each and every document in the document store to identify changes to the document store. (Column 3, lines 63-65). There is no mention or suggestion of "determining whether relevant information is contained in the document," as recited in claims 1, 9 and 17.

In addition, Applicant maintains the argument that Nelson fails to teach or suggest extracting a portion of the document that characterizes the document's subject content to form

the document extract. In Nelson, the compound document, i.e., a document containing multimedia components, is partitioned into its constituent components (e.g., text, audio, images) and one or more tokens is created for each constituent component. No regard is paid to whether the component *characterizes* the document's subject content. For example, the document's subject content might be a particular soccer game, and the document might include images of celebrities attending the soccer game. Nelson's system would extract the images and create a token for the celebrity, although the celebrity has nothing to do with the soccer game. Accordingly, Applicant respectfully submits that none of the references teach or suggest "extracting a portion of the document that characterizes the document's subject content to form the document extract," as recited in claims 1, 9 and 17.

Moreover, none of the references teaches "replacing the document in the storage device with the document extract." In the Final Office Action, the Examiner states that Matsubayashi teaches this feature in the summary and at column 24, lines 19-25. Applicant respectfully disagrees. In the summary, Matsubayashi discusses how single character type strings are extracted from the seed document and from other documents, and how documents related to the seed document are found based on the occurrence frequency of the extracted single character type strings. At column 24, lines 19-25, Matsubayashi states that "the seed document may be replaced by a specified text to similarly extract characteristic strings and to realize the relevant document searching operation." Nothing in Matsubayashi teaches or suggests that "the specified text" is a document extract that *characterizes the document's subject content.* Indeed, "the specified text" can be a portion of the seed document that has no relation to the seed document's subject content, and therefore, under normal operations would have a low occurrence frequency. Accordingly,

Applicant respectfully submits that Matsubayashi fails to teach or suggest "replacing the document in the storage device with the document extract," as recited in claims 1, 9 and 17.

For the reasons presented above, Applicant respectfully submits that the cited references fail to teach or suggest the cooperation of elements recited in claims 1, 9 and 17 and that those claims are therefore allowable over the cited references. Claims 2, 4-8, 10, 12-16 and 18-23 depend on claims 1, 9 and 17, respectively, and the arguments above apply with full force to claims 2, 4-8, 10, 12-16 and 18-23. Accordingly, Applicant respectfully submits that claims 2, 4-8, 10, 12-16 and 18-23 are also allowable over the cited references.

Dependent Claims 5, 6, 13, and 14 are Allowable for Alternative Reasons

Applicant respectfully resubmits that dependent claims 5, 6, 13 and 14 are allowable over the cited references for reasons in addition to being dependent on allowable base claims. First, none of the references teaches or suggests "extracting from the document a collection of sentences that are characteristic of the document's subject content to form a document summary," as recited in claims 5 and 13. In the previous Office Action and Final Office Action, the Examiner states that Nelson teaches this feature at column 5, line 52 to column 6, line 65; column 7, lines 46-67 and column 9, lines 60-65. Those portions, however, discuss tokens and how that are generated. It mentions that "a text component (e.g., a paragraph of text) may be indexed by a number of tokens, each representing one or more words of the text component" (col. 6, lines 10-13), and that "a text token in most cases will represent an actual text string; e.g., the token 'house' will be used to index the word 'house.'" (Col. 6, lines 17-19). Nothing in Nelson teaches or suggests "extracting from the document a collection of sentences *that are characteristic of the document's subject content* to form a document summary," as recited in claims 5 and 13.

Second, none of the references teaches or suggests "selecting from the document extract one of a whole sentence, a portion of a sentence, a word, and a feature," as recited in claims 6 and 14. As discussed above, none of the references teaches or suggests generating the document extract. Therefore, it follows that none of the references can teach or suggest selecting any portion or part of the document extract. In the previous Office Action and the Final Office Action, the Examiner states that Nelson teaches this feature at column 6, lines 16-34, column 7, lines 46-67 and column 9, lines 60-65. Nevertheless, as discussed above, Applicant respectfully submits that the cited portions make no mention or suggestion of "selecting from the document extract one of a whole sentence, a portion of a sentence, a word, and a feature," as recited in claims 6 and 14.

Conclusion

In view of the foregoing, Applicant submits that claims 1, 2, 4-10 and 12-23 are allowable over the cited references. Applicant respectfully requests reconsideration and allowance of the claims as now presented.

Applicant's attorney believes that this application is in condition for allowance. Should any unresolved issues remain, Examiner is invited to call Applicant's attorney at the telephone number indicated below.

Respectfully submitted,

SAWYER LAW GROUP LLP

October 11, 2005
Date

/Joyce Tom/ Reg. No. 48,681
Joyce Tom
Attorney for Applicant(s)
Reg. No. 48, 681
(650) 493-4540